

DEEP LEARNING-BASED DETECTION OF CARDIAC HYPERTROPHY SYMPTOMS IN CT MEDICAL IMAGING

A. Bakunovich¹⁾, S. J. Wang¹⁾

¹⁾ *International Sakharov Environmental Institute of Belarusian State University,
Dolgobrodskaya str., 23/1, 220070, Minsk, Belarus, a.bakunovich@iseu.by*

With advancements in bioinformatics and medical imaging, deep learning-based diagnostic methods are gaining clinical attention. This study proposes an enhanced YOLO11 model with the Convolutional Block Attention Module (CBAM) to improve cardiac hypertrophy detection in chest CT images. By reinforcing channel and spatial attention, the method enhances subtle morphological change detection while reducing interference from soft-tissue densities, blurred boundaries, and anatomical variability. Tests on a public CT dataset show that the improved YOLO11 outperforms the original, achieving 98.19 % precision, 98.79 % recall, and 98.47 % mAP@0.5.

Keywords: *bioinformatics; medical imaging; deep learning; YOLO11.*

ВЫЯВЛЕНИЕ СИМПТОМОВ ГИПЕРТРОФИИ СЕРДЦА НА КТ-СНИМКАХ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ ГЛУБОКОГО ОБУЧЕНИЯ

А. В. Бакунович¹⁾, Ш. Ван¹⁾

¹⁾ *Учреждение образования «Международный государственный экологический институт имени А. Д. Сахарова» Белорусского государственного университета, ул. Долгобродская, 23/1, 220070, г. Минск, Беларусь, a.bakunovich@iseu.by*

С развитием биоинформатики и медицинской визуализации методы диагностики на основе глубокого обучения привлекают все больше внимания в клинической практике. Это исследование предлагает усовершенствованную модель YOLO11 с модулем внимания сверточных блоков (CBAM) для улучшения выявления гипертрофии сердца на КТ-снимках грудной клетки. Усиление канального и пространственного внимания позволяет лучше обнаруживать тонкие морфологические изменения, снижая влияние помех, связанных с плотностью мягких тканей, размытыми границами и анатомическими вариациями. Тестирование на общедоступном наборе КТ-данных показало, что улучшенная YOLO11 превосходит оригинальную модель, достигая 98,19 % точности, 98,79 % полноты и 98,47 % mAP@0.5.

Ключевые слова: *биоинформатика; медицинская визуализация; глубокое обучение; YOLO11.*

<https://doi.org/10.46646/SAKH-2025-1-277-280>

With the rapid evolution of bioinformatics and medical imaging technology, Thoracic CT imaging plays a crucial role in diagnosing cardiac hypertrophy, but detection remains challenging due to low-contrast heart boundaries, anatomical variability, and imaging artifacts. Traditional methods relying on visual inspection and cardiothoracic ratios often lead to diagnostic inaccuracies. Deep learning, particularly CNN-based models, has improved medical image analysis by automating feature extraction. The YOLO series, especially YOLO11, has gained clinical attention due to its real-time detection capabilities. However, its Adaptive Attention Module (AAM) struggles with detecting subtle morphological changes in the heart [1].

YOLO11, released by Ultralytics on September 30, 2024 [1], features high detection accuracy, inference speed, and model efficiency, making it well suited for medical image analysis. The model employs an improved Cross Stage Partial Network (CSPNet) as its backbone, integrating Feature Pyramid Networks (FPN) and Path Aggregation Networks (PAN) for multi-scale feature extraction.

Additionally, GhostNet and RepVGG enhance computational efficiency, while self-distillation learning improves model generalization. The built-in Adaptive Attention Module (AAM) in YOLO11 enhances the detection of small objects and complex backgrounds but remains suboptimal for segmenting large anatomical structures such as the heart in chest CT scans. It is susceptible to interference from surrounding anatomical features, limiting its ability to accurately capture subtle structural variations.

Convolutional Block Attention Module (CBAM) is a lightweight attention mechanism that improves feature representation by predicting attention weights across channel and spatial dimensions and integrating them with the original feature maps [2]. It consists of Channel Attention and Spatial Attention, which enhance model performance with minimal parameter overhead. CBAM emphasizes key features and suppresses irrelevant details by aggregating channel and spatial information, refining intermediate feature maps, and improving attention to salient regions. It is highly adaptable and can be seamlessly integrated into various convolutional network architectures. Unlike AAM, CBAM improves boundary sensitivity and feature extraction for static cardiac structures [2]. This hybrid approach maintains real-time efficiency while significantly improving detection accuracy, demonstrating strong potential for clinical application.

CBAM's channel attention mechanism learns to identify critical feature channels and generates a channel attention map, emphasizing important patterns in the input image. First, the input feature map undergoes max pooling and average pooling along the spatial dimension, producing two complementary spatial context vectors, F_{avg}^c and F_{max}^c . These vectors are then fed into a shared network consisting of a hidden layer and a multilayer perceptron (MLP), which generates the channel attention map M_c . To reduce memory consumption, the hidden layer's activation size is set to $\frac{1}{r}$, where r is the reduction ratio [2]. The max-pooled and average-pooled features processed separately by the shared network and merged via element-wise summation to obtain the final channel attention vector. The calculation method of channel attention is [2]:

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) = \sigma(W_1(W_0(F_{avg}^c)) + W_1(W_0(F_{max}^c)))$$

Here, σ denotes the Sigmoid function, and $\cdot$. Note that the MLP weights W_0 and W_1 are shared, and a ReLU activation follows W_0 .

CBAM's spatial attention mechanism captures spatial relationships in feature maps to identify key regions, complementing channel attention. Specifically, max pooling and average pooling are applied along the channel axis to generate two complementary 2D feature maps, F_{avg}^s and F_{max}^s . These maps are then concatenated and passed through a standard convolution layer to produce the final 2D spatial attention map M_s , which highlights key regions and suppresses irrelevant features. Spatial attention computed as [2]:

$$M_s(F) = \sigma(f^{7 \times 7}([AvgPool(F); MaxPool(F)])) = \sigma(f^{7 \times 7}([F_{avg}^s; F_{max}^s]))$$

Here $f^{7 \times 7}$ represents a 7×7 convolution operation, $[\cdot; \cdot]$ denotes concatenation along the channel dimension, M_s is the spatial attention map. σ represents the sigmoid function.

In the improved YOLO11, the CBAM is placed between the neck and the Detect module [1]. The attention module is arranged as follows: given an input image, the two attention mechanisms (channel and spatial) compute complementary attention, focusing on “what” and “where,” respectively.

The experiments were conducted on a high-performance computing platform. The hardware setup included an NVIDIA A100 GPU with 80 GB of memory, an Intel Xeon Platinum series CPU, and 256 GB of RAM. The software environment consisted of Ubuntu 20.04 LTS as the operating

system, Python 3.9 as the programming language, and PyTorch 1.10.1 with Torchvision 0.11.1 as the deep learning framework. CUDA 10.1 was used for GPU acceleration.

The dataset comprised 1,659 chest CT images, which were preprocessed and standardized. They were divided into a training set of 1,493 images (90 %) and a test set of 166 images (10 %).

To evaluate the model's performance in detecting cardiac hypertrophy from chest CT images, two experiments were conducted over 100 epochs. The first used the original YOLO11 as a baseline to assess detection accuracy without attention mechanisms. The second integrated CBAM into YOLO11 to enhance feature extraction via channel and spatial attention, testing its impact on detection accuracy, sensitivity, and robustness using the same dataset.

The predicted results were classified as follows:

Table 1

Classification of Detection Outcomes

Category	Relevant (Positive Class)	Non-Relevant (Negative Class)
Retrieved (Detected)	True Positives (TP, positive class correctly identified)	False Positives (FP, negative class misclassified as positive)
Not Retrieved (Not Detected)	False Negatives (FN, positive class misclassified as negative)	True Negatives (TN, negative class correctly identified)

In cardiac hypertrophy detection, each bounding box is assigned N+1 confidence scores (N object classes + 1 background). A threshold of 0.5 is set, where detections with confidence ≥ 0.5 are considered positive, otherwise negative.

A higher threshold increases precision but lowers recall, while a lower threshold does the opposite, leading to more false positives. This study adopts 0.5 to balance both metrics.

To evaluate the performance of the proposed model in detecting cardiac hypertrophy, we conducted comparative experiments using the YOLO11 baseline model and YOLO11 with CBAM. The detection results are summarized in Table 2.

Table 2

Performance Comparison Between YOLO11 Baseline and YOLO11 + CBAM

Metric	YOLO11 Baseline	YOLO11 + CBAM
Precision	96.21 %	98.19 %
Recall	96.97 %	98.79 %
mAP@0.5	97.11 %	98.47 %
mAP@0.95	55.12 %	58.65 %

Precision represents the number of actually positive samples among the predicted samples divided by the total number of positive samples. The formula is:

$$Precision = \frac{TP}{TP + FP}$$

Recall represents the number of actually positive samples among the predicted samples divided by the total number of predicted samples. The formula is:

$$Recall = \frac{TP}{TP + FN}$$

AP (Average Precision) refers to the average precision, indicating the mean accuracy for all images of a certain class. By plotting Recall on the x-axis and Precision on the y-axis to form the P-R curve, AP is obtained from the area under this curve.

$$AP = \int_0^1 p(r)dr$$

mAP (Mean Average Precision), the main evaluation metric for object detection algorithms, is the mean of Average Precision across different recall levels. A higher mAP value indicates better detection performance on the given dataset. For a test set with N classes, the formula for mAP is:

$$mAP = \frac{\sum_{k=1}^N AP_k}{N}$$

The results demonstrate that integrating CBAM into YOLO11 significantly enhances model performance across multiple metrics. Precision increased from 96.21% to 98.19%, reducing false positives and improving specificity. Recall rose from 96.97% to 98.79%, reflecting better detection of cardiac hypertrophy cases. mAP@0.5 improved from 97.11% to 98.47%, ensuring more accurate localization of bounding boxes. mAP@0.5 represents the mean Average Precision (mAP) at an Intersection over Union (IoU) threshold of 0.5, meaning that for each detection, the model's predicted box and the ground truth box need to have an IoU greater than or equal to 0.5 to be considered a correct detection. Meanwhile, mAP@0.95 increased from 55.12% to 58.65%, indicating greater robustness across different IoU thresholds. mAP@0.95 represents the mean AP at an IoU threshold of 0.95, where the predicted and ground truth boxes need to have an IoU greater than or equal to 0.95 to be considered a correct detection. This improvement in both mAP@0.5 and mAP@0.95 is particularly evident in refining heart structure boundaries. IoU is calculated as the area of intersection between the predicted and ground truth boxes divided by the area of their union.

These findings confirm that YOLO11+CBAM outperforms the baseline in classification and localization accuracy. CBAM enhances the model's ability to focus on relevant features, reducing background noise and improving feature extraction, which is crucial for detecting cardiac hypertrophy. Through rigorous validation, the enhanced model consistently exceeded baseline results across all metrics.

This study underscores the importance of attention mechanisms in medical image analysis. By emphasizing subtle changes in the cardiac region, CBAM significantly boosts detection accuracy, maintaining high performance even under stringent conditions. Ultimately, this research highlights deep learning's potential in bioinformatics and medical imaging, providing valuable insights for future applications of attention mechanisms in complex medical data analysis.

References

1. YOLOv11: An Overview of the Key Architectural Enhancements [Electronic resource]. URL: <https://arxiv.org/abs/2410.17725> (access date: 06.02.2025).
2. Woo S., Park J., Lee J.-Y., Kweon I. S. CBAM: Convolutional Block Attention Module // Proceedings of the European Conference on Computer Vision (ECCV). 2018. Vol. 11211. P. 3–19.